ISEG - Lisbon School of Economics and Management

List of Exercises - Chapters 6 and 7
1^{st} Semester of 2020/2021

December 19, 2020

1. A box contains 10 balls, of which 3 are red, 2 are yellow, and 5 are blue. Five balls are randomly selected with replacement.

   (a) Calculate the probability that fewer than 2 of the selected balls are red.

   Answer: Let X be the random variable that represents the number of red balls in five trials, assuming that the experiment is done with replacement. Then

   $$X \sim Bin(n = 5, p = 3/10),$$

   where p is indeed

   $$p = P(\text{"a red ball is selected"}) = \frac{3}{10}.$$

   Then, the requested probability is

   $$P(X < 2) = \tbinom{5}{0}\left(\frac{3}{10}\right)^0\left(\frac{7}{10}\right)^5 + \tbinom{5}{1}\left(\frac{3}{10}\right)^1\left(\frac{7}{10}\right)^4 \approx 0.528$$

   (b) Assume now that the five balls are randomly selected without replacement. Compute the previous probability.

   Answer: Let Y be the random variable that represents the number of red balls in five trials, assuming that the experiment is done without replacement. Then

   $$Y \sim Hypergeometric(N = 10, M = 3, n = 5).$$

   Then, the requested probability is

   $$P(X < 2) = \frac{\tbinom{7}{5}\tbinom{3}{0}}{\tbinom{10}{5}} + \frac{\tbinom{7}{4}\tbinom{3}{1}}{\tbinom{10}{5}} = 0.5$$

2. Prove that if $X_1 \sim B(n_1, p)$ and $X_2 \sim B(n_2, p)$ and X_1 and X_2 are independent random variables, then $X_1 + X_2 \sim B(n_1 + n_2, p)$ (**Hint:** Recall that if $X \sim B(n, p)$, then $M_X(t) = [(1-p) + pe^t]^n$.)

3. Let $X \sim B(n, p)$ and $X^* \sim B(n, 1-p)$, show that $P(n - X^* = x) = P(X = x)$.

   Answer: By definition,

$$P(n - X^* = x) = P(X^* = n - x) = \binom{n}{n-x}(1-p)^{n-x}p^x.$$

   Since $\binom{n}{n-x} = \frac{n!}{(n-x)!x!} = \binom{n}{x}$, then

$$P(n - X^* = x) = \binom{n}{x}(1-p)^{n-x}p^x = P(X = x).$$

4. Just prior to jury selection for O. J. Simpson's murder trial in 1995, a poll found that about 20% of the adult population believed Simpson was innocent (after much of the physical evidence in the case had been revealed to the public). Assume that the 12 jurors were selected randomly and independently from the population

   (a) Find the probability that the jury had at least one member who believed in Simpson's innocence.

   Answer: Let X be the random variable that represents the number of jurors, in a sample of 12 jurors, who believed in Simpson's innocence

$$X \sim Bin(n = 12, p = 1/5).$$

   Then, the requested probability is

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \binom{12}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{12} \approx 0.9313$$

   (b) Find the probability that the jury had at least two members who believed in Simpson's innocence.

   Answer:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - P(X = 1) - P(X = 0) = 0.7251$$

   where

$$P(X = 1) = \binom{12}{1}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^{11} \approx 0.2062.$$

   (c) What is the expected value and the variance of the number of jurors, in a sample of 12 jurors, who believed in Simpson's innocence?

   Answer: $E(X) = 12 * 0.2 = 2.4$ and $Var(X) = 12 * 0.8 * 0.2 = 1.92$

5. A student takes a multiple choice test with 20 questions, each with 4 choices and only one is correct.

   (a) Assume that the student blindly guesses and gets one question correct. Find the probability that the reader has to read no more than 4 questions until we get the right one.

   Solution: Let X be the random variable that represents the number of questions that the reader needs to read until we get the right one.

   $$X \sim Geo(p = 1/4)$$

   Then,

   $$P(X \leq 4) = \sum_{x=1}^{4} (1/4)^1 (3/4)^{x-1} = 0.683$$

   (b) Assume that the student blindly guesses and gets two questions correct. Find the probability that the reader has to read 10 questions until he gets the second question that is correct.

   Solution: Let Y be the random variable that represents the number of questions that the reader needs to read until we get the first two right questions.

   $$X \sim NB(k = 2, p = 1/4)$$

   Then,

   $$P(X = 10) = \binom{9}{1} (1/4)^2 (3/4)^8 = 0.0563$$

6. Past Experience indicates that an average number of 6 customers per hour stop for petrol at a petrol station. Assuming that the number of customers that stop for petrol at a petrol station is a Poisson random variable:

   (a) What is the probability of 3 customers stopping in any hour?

   Solution: Let X be a a random variable representing the number of customers that stop for petrol at a petrol station.

   $$X \sim Poi(\lambda), \quad \lambda = E(X) = 6.$$

   Then,

   $$P(X = 3) = \frac{e^{-6}\, 6^3}{3!} \approx 0.089$$

   (b) What is the probability of 3 customers or less in any hour?

   $$P(X \leq 3) = \sum_{x=0}^{3} \frac{e^{-6} 6^x}{x!} \approx 0.1512.$$

(c) What is the expected value, and standard deviation of the distribution.
Since $E(X) = 6$ and $Var(X) = 6$, then $\sigma_X = \sqrt{6}$.

7. The average number of trucks arriving on any one day at a truck depot in a certain city is known to be 2.

(a) If we assume a Poisson distribution, what is the probability that on a given day fewer than 2 trucks will arrive at this depot?

Solution: Let X be random variable that represents the number of trucks arriving on any one hour at a truck depot in a certain city,

$$X \sim Poi(\lambda), \quad \lambda = E(X) = 2.$$

Then,

$$P(X < 2) = P(X = 0) + P(X = 1) = \frac{e^{-2}\, 2^0}{0!} + \frac{e^{-2}\, 2^1}{1!} \approx 0.41.$$

(b) Assume that in a set of 10 random days the number of trucks arriving at the truck depot is independent in each day. Compute the probability of having fewer than 2 trucks arriving at the truck depot each day, for 5 days.

Solution: Let Y be the random variable that represents the number of days in which fewer than 2 trucks will arrive at the truck depot, in a group of 5 days.

$$Y \sim Bin(n = 10, p = 0.41)$$

because $p = P(X < 2) \approx 0.41$. Then the requested probability is

$$P(Y = 5) = \binom{10}{5} (0.41)^5 (1 - 0.41)^5 \approx 0.209$$

8. (Lack of memory of the exponential random variable) Let $X \sim Exp(\lambda)$, prove that $P(X > x + s | X > x) = P(X > s)$ for any $x \geq 0$ and $s \geq 0$.

9. Let $X_i \sim Exp(\lambda_i)$, $i = 1, 2$, be independent random variables. Prove that $Y = \min\{X_1, X_2\} \sim Exp(\lambda_1 + \lambda_2)$.

(**Hint:** Note that $P(Y > y) = P(X_1 > y, X_2 > y)$).

Solution: If $Y = \min\{X_1, X_2\}$, where $X_i \sim Exp(\lambda_i)$, $i = 1, 2$, then

$$\begin{aligned} 1 - F_Y(y) = P(Y > y) &= P(X_1 > y, X_2 > y) = P(X_1 > y) \times P(X_2 > y) \\ &= (1 - F_{X_1}(y))(1 - F_{X_2}(y)) = e^{-\lambda_1 y} e^{-\lambda_2 y} = e^{-(\lambda_1 + \lambda_2) y}. \end{aligned}$$

Therefore, $Y \sim Exp(\lambda_1 + \lambda_2)$.

10. Let $X_i \sim Exp(\lambda)$, $i = 1, 2$, be independent random variables.

   (a) Find the distribution of $Y_1 = \min\{X_1, X_2\}$ and $Y_2 = \max\{X_1, X_2\}$.
   Solution: From the previous exercise,

$$Y_1 \sim Exp(2\lambda).$$

   To derive the distribution of Y_2, we can see that

$$\begin{aligned} F_{Y_2}(y) = P(Y_2 < y) = P(X_1 < y, X_2 < y) &= P(X_1 < y) \times P(X_2 < y) \\ &= F_{X_1}(y) F_{X_2}(y) = \left(1 - e^{-\lambda y}\right)\left(1 - e^{-\lambda y}\right) \\ &= 1 - 2e^{-\lambda y} + e^{-2\lambda y}, \quad y > 0. \end{aligned}$$

   One can see that,

$$f_{Y_2}(y) = \begin{cases} 2\lambda e^{-\lambda y} - 2\lambda e^{-2\lambda y}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

   (b) Find the expected value of $Z = \lambda^2 Y_1 - \frac{2}{3} Y_2$.
   By definition,

$$E(Y_2) = \int_{-\infty}^{\infty} y f_{Y_2}(y) dy = \frac{3}{2}\lambda,$$

   therefore,

$$E(Z) = E\left(\lambda^2 Y_1 - \frac{2}{3} Y_2\right) = \frac{\lambda}{2} - \frac{1}{\lambda}$$

11. The lifetime in years of an electronic component is a continuous random variable X that follows

$$X \sim Exp(1)$$

   (a) Find the lifetime L which a typical component is 60% certain to exceed.
   Solution: Taking into account that X is an exponential with parameter 1, we have that

$$0.6 = P(X > L) = 1 - F_X(L) \Leftrightarrow e^{-L} = 0.6 \Leftrightarrow L \approx 0.51.$$

   Therefore, the lifetime of 60% of that electronic components exceed 0.51 hours.

   (b) If five components are sold to a manufacturer, find the probability that at least one of them will have a lifetime less than L years.

Solution: Let Y be the random variable that represents the number of components, in a set of five, with a lifetime less than L years.

$$Y \sim Bin(n = 5, p),$$

with $p = P(X \leq L) = 1 - P(X > L) = 0.4$. Therefore, the required probability is

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) = 1 - 0.6^5 = 0.92.$$

12. The time intervals between successive trains stopping in a certain rail station have an exponential distribution with mean 6 minutes.

    (a) Find the probability that the time interval between two consecutive trains is less than 5 minutes.

    Answer: 0.57

    (b) Find a time interval t such that we can be 95% sure that the time interval between two successive trains will be greater than t.

    Answer: 0.308

    (c) Assume that the number of trains arriving in one hour is modeled by a Poisson random variable. Compute the probability that in a random hour 5 trains stop at the train station.

    Answer: 0.04 (Notice that, in hours, the expected time interval between trains is = 0.1.)

    (d) If we have counted 8 trains in the first hour, what is the probability that two of them arrived in the first 30 minutes?

    Solution: 0.1094

13. Compute the following probabilities:

    (a) If Y is distributed $\chi^2(4)$ find $P(Y \leq 7.78)$.

    Answer: 0.9

    (b) If Y is distributed $\chi^2(10)$ find $P(Y > 18.31)$.

    Answer: 0.05

    (c) If Y is $\chi^2(1)$ find $P(Y \leq 3.8416)$.

    Answer: 0.95

14. Using the moment generating function, show that if $X \sim Gamma(a, b)$ and $Y = 2X/b$, then $Y \sim \chi^2(2a)$.

15. Prove that if X_1 and X_2 are independent random variables with Gamma distribution $X_1 \sim Gamma(a_1, b)$ and $X_2 \sim Gamma(a_2, b)$, then $X_1 + X_2 \sim Gamma(a_1 + a_2, b)$. (**Hint:** Recall that if $X \sim Gamma(a, b)$, then $M_X(t) = (1 - bt)^{-a}$ for $t < 1/b$).

    Solution: Assume that $Y = X_1 + X_2$, then the moment generating function is given by

$$\begin{aligned} M_Y(t) = E(e^{tY}) = E(e^{t(X_1+X_2)}) &= \underbrace{E(e^{tX_1})E(e^{tX_2})}_{X_1 \text{ and } X_2 \text{ are independent}} \\ &= (1-bt)^{-a_1}(1-bt)^{-a_2} = (1-bt)^{-(a_1+a_2)}, \quad \text{for all } t < 1/b, \end{aligned}$$

    which means that $Y \sim Gamma(a_1 + a_2, b)$.

16. Suppose customers arrive at a store according to a Poisson process, where the expected number of customers per hour is 0.5.

    (a) Knowing that 4 customers arrived at the store during the morning (4 hours) what is the probability that in this day (8 hours) the store receives more than 15 customers?

    Answer: Let X_t be the random variable that counts the number of customers arrive at a store.

$$X_t \sim Poisson(t\lambda), \quad \lambda = \frac{1}{2}$$

    The required probability is

$$\begin{aligned} P(X_8 > 15 | X_4 = 4) &= \frac{P(X_8 > 15,\ X_4 = 4)}{P(X_4 = 4)} = \frac{P(X_8 - X_4 > 15 - 4)P(X_4 = 4)}{P(X_4 = 4)} \\ &= P(X_4 > 11) = 1 - F_X(11) \end{aligned}$$

    (b) Compute the probability that the first customer does not arrive during the first hour (since the opening hour of the store).

    Answer: Let Z be the random variable that represents the time spent until the first customer arrives.

$$Z \sim Exp(0.5).$$

    The required probability is:

$$P(Z > 1) = 1 - F_Z(1) = e^{-1/2} = 0.61.$$

    (c) What is the distribution of the time until the second customer arrives?

    Answer: Let X be the time spent until the second customer arrives. Then X is such that

$$X \sim Gamma(2, 1/0.5) = Gamma(2, 2) = \chi^2_{(4)}.$$

(d) Find the probability that one has to wait at least half an hour until the second customer arrives.

Answer: By using the tables regarding the χ^2 distribution, we get that

$$P(X \geq 0.5) \approx 0.975$$

(e) Find the probability that one has to wait at least five hours until the fourth customer arrives.

Let Y be the time spent until the fourth customer arrives. Then Y is such that

$$Y \sim Gamma(4, 1/0.5) = Gamma(4, 2) = \chi^2_{(8)}.$$

The required probability is

$$P(Y \geq 5) \approx 0.75.$$

17. Compute the following probabilities:

(a) If Y is distributed $N(1, 4)$, find $P(Y \leq 3)$.

Answer: 0.841

(b) If Y is distributed $N(3, 9)$, find $P(Y > 0)$.

Answer: 0.841

(c) If Y is distributed $N(50, 25)$, find $P(40 \leq Y \leq 52)$.

Answer: 0.633

(d) If Y is distributed $N(0, 1)$, find $P(|Y| > 1.96)$.

Answer: 0.05

18. Prove that if the random variables $X_i, i = 1, 2$ have a normal distribution, $X_i \sim N(\mu_i, \sigma_i^2)$, and are independent and if $Y = aX_1 + bX_2 + c$, then $Y \sim N(\mu_Y, \sigma_Y^2)$, where $\mu_Y = a\mu_1 + b\mu_2 + c$ and $\sigma_Y^2 = a^2\sigma_1^2 + b^2\sigma_2^2$.

(**Hint:** Recall that if $X \sim N(\mu, \sigma^2)$, then $M_X(t) = e^{\left(\mu t + 0.5\sigma^2 t^2\right)}$ and note that functions of independent random variables are also independent).

19. Suppose that diameter of a certain component produced in a factory can be modeled by a normal distribution with mean $10cm$ and standard deviation $3cm$.

(a) Find the probability that the diameter of a random component is larger than $13cm$.

Answer: Let X be the diameter of a certain component produced in a factory.

$$X \sim N(\mu = 10, \sigma^2 = 9).$$

Then,

$$P(X > 13) = P\left(\frac{X-10}{3} > \frac{13-10}{3}\right) = P(Z > 1) = 1 - \Phi(1) = 0.159$$

where $Z \sim N(0,1)$.

(b) Find the probability that the diameter of a random component is less than $7cm$.

Answer:

$$P(X < 7) = P\left(\frac{X-10}{3} < \frac{7-10}{3}\right) = P(Z < -1) = 1 - \Phi(1) = 0.159$$

the last equality following in light of the symmetry of the normal distribution.

(c) Selecting randomly 10 components, what is the probability that 2 of them have a diameter less than $7cm$?

Answer: Let Y be the random variable that represents the number of components that have a diameter less than $7cm$, in a set of 10 components.

$$Y \sim Bin(n = 10, p = 0.159)$$

because $p = P(X < 7)$. The requested probability is

$$P(Y = 2) = \binom{10}{2} 0.159^2 (1 - 0.159)^8 = 0.285.$$

(d) What is the expected number of components that we have to inspect to find 1 with a diameter less than $7cm$?

Answer: Let Z be the random variable that we have to inspect to find 1 with a diameter less than $7cm$.

$$Z \sim Geo(p), \quad \text{where } p = P(X < 7) = 0.159$$

Then,

$$E(Z) = 1/p \approx 6.3,$$

which means that, in average, one has to inspect 7 components.

20. A baker knows that the daily demand for a specific type of bread is a random variable X such that $X \sim N(\mu = 50, \sigma^2 = 25)$. Find the demand which has probability 1% of being exceeded.

Answer: 61.65

21. Assume that X_i, with $i = 1, 2, 3$ represent the profit, in million of euros, of 3 different companies located in 3 different countries. If

$$X_1 \sim N(1, 0.01), \quad X_2 \sim N(1.5, 0.02), \quad X_3 \sim N(2, 0.06)$$

(a) Which company is more likely to have a profit greater than 1.5 millions?

Answer: Due to the symmetry of the normal distributions, we know that

$$P(X_1 > 1.5) < P(X > 1) = 0.5, \quad P(X_2 > 1.5) = 0.5$$
$$P(X_3 > 1.5) > P(X_3 > 2) = 0.5.$$

Therefore, company 3 is more likely to exceed a profit of 1.5 million.

(b) What is the probability of the profit of these 3 companies does not exceed 4 millions of euros? (Assume independence.)

From the properties of independent normal random variables, we know that

$$X_1 + X_2 + X_3 \sim N(\mu = 1 + 1.5 + 2, \sigma^2 = 0.01 + 0.02 + 0.06).$$

Therefore, if $Z \sim N(0, 1)$, then

$$P(X_1 + X_2 + X_3 < 4) = P\left(\frac{X_1 + X_2 + X_3 - 4.5}{0.3} < \frac{4 - 4.5}{0.3}\right)$$
$$= P(Z < -5/3) \approx 0.05.$$

22. The time elapsed since failure until repair (designated as repair time) of a certain type of machines is a random variable with exponential distribution with mean of 2 hours.

(a) What is the probability that a broken machine has a repair time of 1 hour or less?

Answer: Let X be the random variable that represents the time (in hours) elapsed since failure until repair.

$$X \sim Exp(\lambda), \quad \lambda = 1/E(X) = 0.5.$$

Then,

$$P(X \leq 1) = F_X(1) = 1 - e^{-1\times 0.5} = 0.393.$$

(b) If 10 broken machines were randomly selected, what is the probability of the fastest repair be performed in less than 15 minutes? (assume independence)

Answer: Let X_i, with $i = 1 \cdots, 10$, be the random variable that represents the time (in hours) elapsed since failure until repair of the i^{th} machine.

$$X_i \sim Exp(\lambda), \quad \lambda = 1/E(X) = 0.5.$$

and the the random variables $X_1, \cdots, X_{10}$ are independent. From what we already saw $\min\{X_1, \cdots, X_{10}\} \sim Exp(5)$. Then,

$$P(\min\{X_1, \cdots, X_{10}\} < 1/4) = 1 - e^{-5/4} \approx 0.713$$

(c) What is the probability that the total repair time of 50 broken machines does not exceed 90 hours? (assume independence)

Answer: Let X_i, with $i = 1 \cdots, 50$, be the random variable that represents the time (in hours) elapsed since failure until repair of the i^{th} machine.

$$X_i \sim Exp(\lambda), \quad \lambda = 1/E(X) = 0.5.$$

and the the random variables $X_1, \cdots, X_{50}$ are independent. Then the total repair time of 50 broken machines is given by

$$T_{50} = \sum_{i=1}^{50} X_i.$$

Due to the Limit Central Theorem, we get that

$$T_{50} = \sum_{i=1}^{50} X_i \overset{a}{\sim} N(\mu, \sigma^2)$$

where

$$\mu = \sum_{i=1}^{50} E(X_i) = \sum_{i=1}^{50} \frac{1}{\lambda} = 100$$

and

$$\sigma^2 = \sum_{i=1}^{50} Var(X_i) = \sum_{i=1}^{50} \frac{1}{\lambda^2} = 200.$$

Then,

$$P(T_{50} < 90) = P\left(\frac{T_{50} - 100}{10\sqrt{2}} < \frac{90 - 100}{10\sqrt{2}}\right) \approx P(Z < -0.71) = 0.239.$$

23. Suppose that you roll a balanced die 36 times. Let Y denote the sum of the outcomes in each of the 36 rolls. Estimate the probability that $108 \leq Y \leq 144$.

    Solution: 0.858

24. Suppose that a book with 300 pages contains on average 1 misprint per page. Assume that the number of misprints per page is a Poisson random variable.

    (a) What is the probability that a random page has 2 or more misprints?
    Solution: 0.264

    (b) What is the probability that there will be at least 100 pages which contain 2 or more misprints? (assume independence)
    Solution: 0.0033

(c) What is the probability that there will be no more than 200 misprints in the book?

Solution: ≈ 0

25. Assume that the number of hours per week that a student spends studying for the course of Statistics 1 follows a continuous uniform distribution in the interval $(0, 5)$.

(a) What is the probability that a random student spends more than 3 hours studying for the course of Statistics 1?

Solution: Let X be the random variable that represents the number of hours that students spend studying for the course of Statistics 1.

$$X \sim U(0, 5).$$

Then,

$$P(X > 3) = \int_3^5 \frac{3}{5} dx = \frac{2}{5}.$$

(b) In a group of 300 students, what is the probability that more than 100 spend more than 3 hours studying for the course of Statistics 1?

Solution: Let Y be the random variable that represents the number of students, in 300, that spend more than 3 hours studying for the course of Statistics 1.

$$Y \sim Bin(n = 300, p), \quad \text{with } p = P(X \geq 3) = \frac{2}{5}.$$

As the number of trials is large enough, the central limit theorem allows us to say that

$$Y \overset{a}{\sim} N(\mu, \sigma^2)$$

where,

$$\mu = n \times p = 120 \quad \text{and} \quad \sigma^{=} n \times p \times (1 - p) = 72.$$

Therefore,

$$P(Y > 100) = P\left(\frac{Y - 120}{\sqrt{72}} > \frac{100 - 120}{\sqrt{72}}\right) \approx P(Z > -2.36) = 0.99$$

(c) In a group of 300 students, what is the probability that, on average, students spend more than 4 hours studying for the course of Statistics 1?

Solution: Let X_i, with $i = 1, \cdots, 300$, be the random variable that represents the number of hours that student i spends studying for the course of Statistics 1. Then

$$X_i \sim U(0, 5), \quad \text{for } i = 1, \cdots, 300$$

and X_i, with $i = 1, \cdots, 300$ are independent random variables. Therefore, the average number of hours spent by students studying for the course of statistics one is modeled by

$$\overline{X} = \frac{1}{300} \sum_{i=1}^{300} X_i.$$

From the properties of expected value and variance, we get

$$E(\overline{X}) = \frac{1}{300} \sum_{i=1}^{300} E(X_i) = E(X_i) = 2,5$$

and

$$Var(\overline{X}) = \underbrace{\left(\frac{1}{300}\right)^2 \sum_{i=1}^{300} Var(X_i)}_{\text{due to independence}} = \frac{1}{300} Var(X_i) = \frac{1}{300} \times \frac{25}{12}.$$

Therefore, from the central limit theorem, we get that

$$Z = \frac{\overline{X} - 2,5}{\sqrt{\frac{1}{300} \times \frac{25}{12}}} \overset{a}{\sim} N(0,1).$$

The intended probability follows

$$P(\overline{X} > 4) \approx P\left(Z > \frac{4 - 2,5}{\sqrt{\frac{1}{300} \times \frac{25}{12}}}\right) = P(Z > 18) \approx 0.$$